WHAT IS CLAIMED IS:

1. A method for normalizing a score associated with a document comprising the steps, performed by a processor, of:

(a) determining statistics relating to scores assigned to a set of training documents not relevant to a topic, the scores representing a measure of relevance to the topic;

(b) normalizing a score assigned to a testing document based on the statistics;

(c) comparing the normalized score to a threshold score; and

(d) designating the testing document as relevant or not relevant to the topic based on the comparison.

2. The method of claim 1, wherein the statistics include a mean score of the training documents not relevant to the topic and a standard deviation of the scores assigned to the set of training documents not relevant to the topic.

3. The method of claim 2, wherein said normalizing step determines the normalized score according the following formula:

$$normalized\_score = (s - \mu_{off\_topic}) / \sigma_{off\_topic}$$

wherein s represents the score assigned to the testing document, μ_{off_topic} represents the mean score of the documents not relevant to the topic, and σ_{off_topic} represents the standard deviation of the scores assigned to the set of training documents not relevant to the topic.

4. The method of claim 1, said determining step further comprising:

determining statistics relating to scores assigned to a set of training documents relevant to the topic.

5. The method of claim 4, wherein the statistics relating to scores assigned to a set of training documents relevant to the topic include a mean score of the documents relevant to the topic and a standard deviation of the scores assigned to the set of training documents relevant to the topic.

6. The method of claim 5, wherein said normalizing step comprises:

normalizing a score assigned to a testing document based on the statistics relating to the scores assigned to the set of training documents not relevant to the topic and based on the statistics relating to the scores assigned to the set of training documents relevant to the topic.

7. The method of claim 6, wherein said normalizing step determines the normalized score according the following formula:

$$\text{normalized\_score} = f_{\text{on-topic}} * ((s - \mu_{\text{off\_topic}}) / \sigma_{\text{off\_topic}})$$

wherein $f_{\text{on-topic}}$ represents a scale factor based on the statistics relating to the scores assigned to the set of training documents relevant to the topic, s represents the score assigned to the testing document, $\mu_{\text{off\_topic}}$ represents the mean score of the documents not relevant to the

topic, and σ_{off_topic} represents the standard deviation of the scores assigned to the set of training documents not relevant to the topic.

8. The method of claim 1, wherein said designating step comprises:

designating the testing document as relevant to the topic based on a determination that the normalized score is greater than the threshold score; and

designating the testing document as not relevant to the topic based on a determination that the normalized score is not greater than the threshold score.

9. The method of claim 1, further comprising:

repeating steps (a)-(d) for a plurality of topics.

10. The method of claim 1, further comprising:

repeating steps (a)-(d) for a plurality of testing documents.

11. The method of claim 1, wherein the statistics include a robust estimate of a mean score of the training documents not relevant to the topic and a robust estimate of a standard deviation of the scores assigned to the set of training documents not relevant to the topic.

12. A data processing system for normalizing a score associated with a document, comprising:

a memory having program instructions; and

a processor responsive to the program instructions to determine statistics relating to scores assigned to a set of training documents not relevant to a topic, the scores representing a measure of relevance to the topic; normalize a score assigned to a testing document based on the statistics; compare the normalized score to a threshold score; and designate the testing document as relevant or not relevant to the topic based on the comparison.

13. A computer-readable medium containing instructions for performing a method for normalizing a score associated with a document, the method comprising:

(a) determining statistics relating to scores assigned to a set of training documents not relevant to a topic, the scores representing a measure of relevance to the topic;

(b) normalizing a score assigned to a testing document based on the statistics;

(c) comparing the normalized score to a threshold score; and

(d) designating the testing document as relevant or not relevant to the topic based on the comparison.

14. The computer-readable medium of claim 13, wherein the statistics include a mean score of the training documents not relevant to the topic and a standard deviation of the scores assigned to the set of training documents not relevant to the topic.

15. The computer-readable medium of claim 14, wherein said normalizing step

determines the normalized score according the following formula:

$$normalized\_score = (s - \mu_{off\_topic}) / \sigma_{off\_topic}$$

wherein s represents the score assigned to the testing document, μ_{off_topic} represents the mean score of the documents not relevant to the topic, and σ_{off_topic} represents the standard deviation of the scores assigned to the set of training documents not relevant to the topic.

16. The computer-readable medium of claim 13, said determining step further comprising:

determining statistics relating to scores assigned to a set of training documents relevant to the topic.

17. The computer-readable medium of claim 16, wherein the statistics relating to scores assigned to a set of training documents relevant to the topic include a mean score of the documents relevant to the topic and a standard deviation of the scores assigned to the set of training documents relevant to the topic.

18. The computer-readable medium of claim 17, wherein said normalizing step comprises:

normalizing a score assigned to a testing document based on the statistics relating to the scores assigned to the set of training documents not relevant to the topic and based on the statistics relating to the scores assigned to the set of training documents relevant to the topic.

19. The computer-readable medium of claim 18, wherein said normalizing step determines the normalized score according the following formula:

normalized_score = $f_{\text{on-topic}}$ * ((s - $\mu_{\text{off\_topic}}$) / $\sigma_{\text{off\_topic}}$)

wherein $f_{\text{on-topic}}$ represents a scale factor based on the statistics relating to the scores assigned to the set of training documents relevant to the topic, s represents the score assigned to the testing document, $\mu_{\text{off\_topic}}$ represents the mean score of the documents not relevant to the topic, and $\sigma_{\text{off\_topic}}$ represents the standard deviation of the scores assigned to the set of training documents not relevant to the topic.

20. The computer-readable medium of claim 13, wherein said designating step comprises:

designating the testing document as relevant to the topic based on a determination that the normalized score is greater than the threshold score; and

designating the testing document as not relevant to the topic based on a determination that the normalized score is not greater than the threshold score.

21. The computer-readable medium of claim 13, further comprising:

repeating steps (a)-(d) for a plurality of topics.

22. The computer-readable medium of claim 13, further comprising:

repeating steps (a)-(d) for a plurality of testing documents.

23. The computer-readable medium of claim 13, wherein the statistics include a robust estimate of a mean score of the training documents not relevant to the topic and a robust estimate of a standard deviation of the scores assigned to the set of training documents not relevant to the topic.

24. A method for normalizing a score associated with a document comprising the steps, performed by a processor, of:

receiving a query including a topic;

determining statistics relating to scores assigned to a set of training documents not relevant to the topic, the scores representing a measure of relevance to the topic; and

normalizing a score assigned to a testing document based on the statistics.

25. The method of claim 24, further comprising:

designating the testing document as relevant or not relevant to the topic based on the normalized score.

26. The method of claim 24, further comprising:

comparing the normalized score to a threshold score; and

designating the testing document as relevant or not relevant to the topic based on the

comparison.

27. A method for searching for documents relevant to a topic comprising the steps, performed by a processor, of:

sending a query including a topic to a processor; and

receiving results from the processor indicating a document relevant to the topic, wherein the processor determines statistics relating to scores assigned to a set of training documents not relevant to a topic, normalizes a score assigned to a testing document based on the statistics, and designates the testing document as relevant or not relevant to the topic based on the normalized score.